

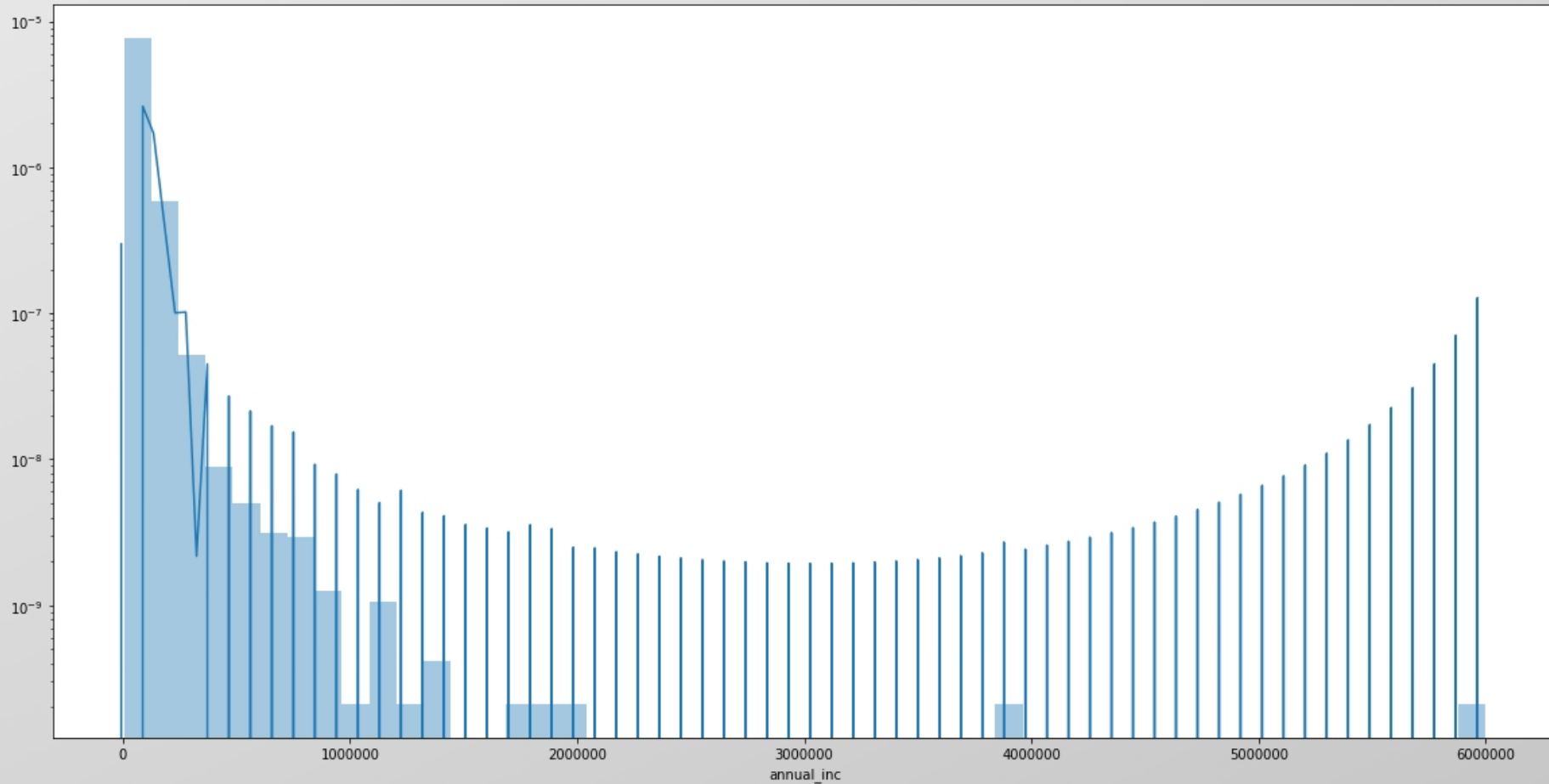
EDA case study

Case study objective:
the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Data Understanding:

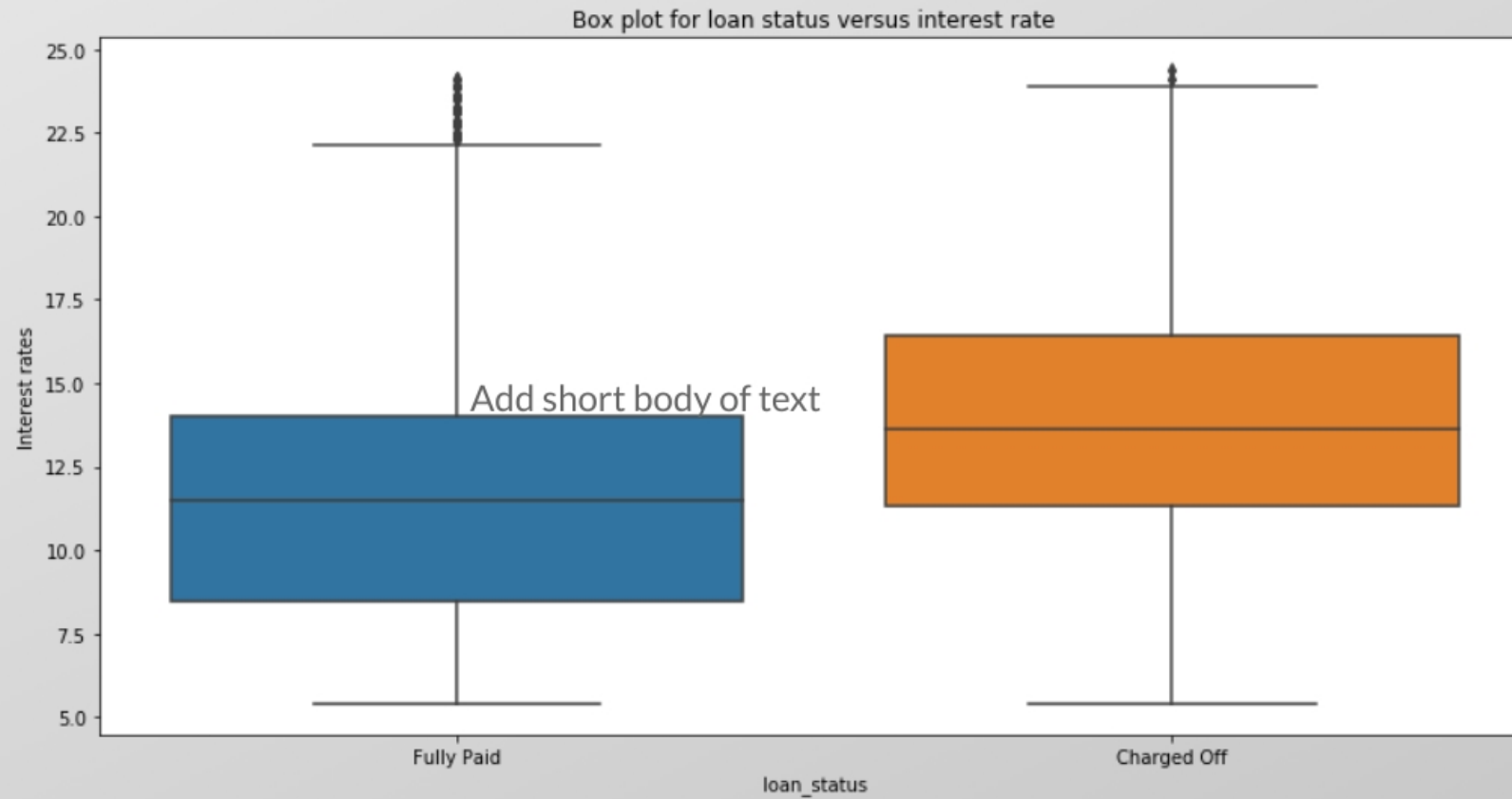
From the loan data set obtain all loan issued through the time period 2007-2011

Performing univariate analysis of annual income column



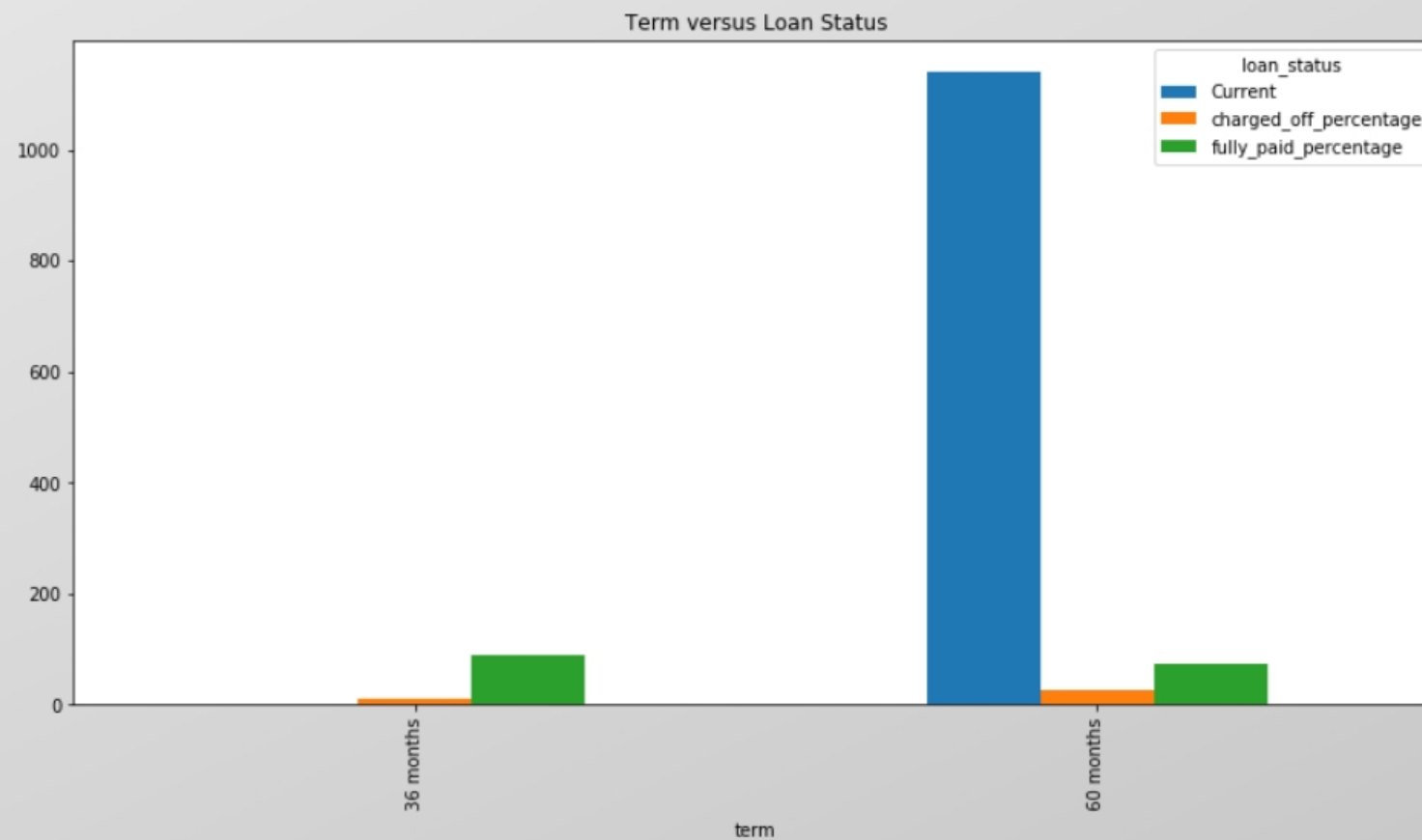
From the above plot we can see that majority of loan applicant's annual income is less than 10 lakhs. Outlier detection: From the above plot we can observe that there are some outliers. But because of very few outliers, mean has not been skewed.

Bivariate analysis on loan status and interest rate



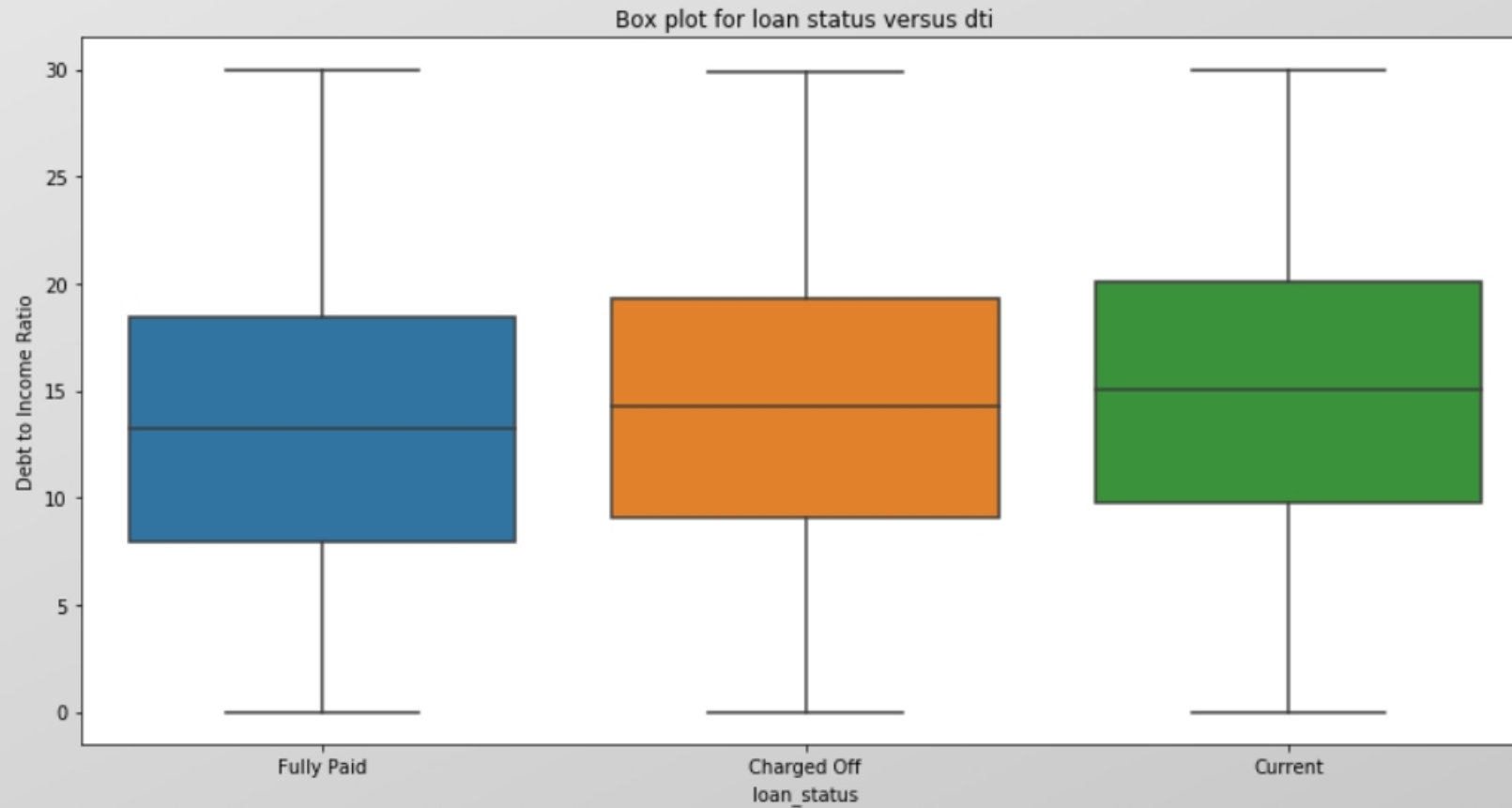
We can see from the above plot that higher the interest rate, higher the chances of loan getting charged off

Bivariate analysis on loan status and Term



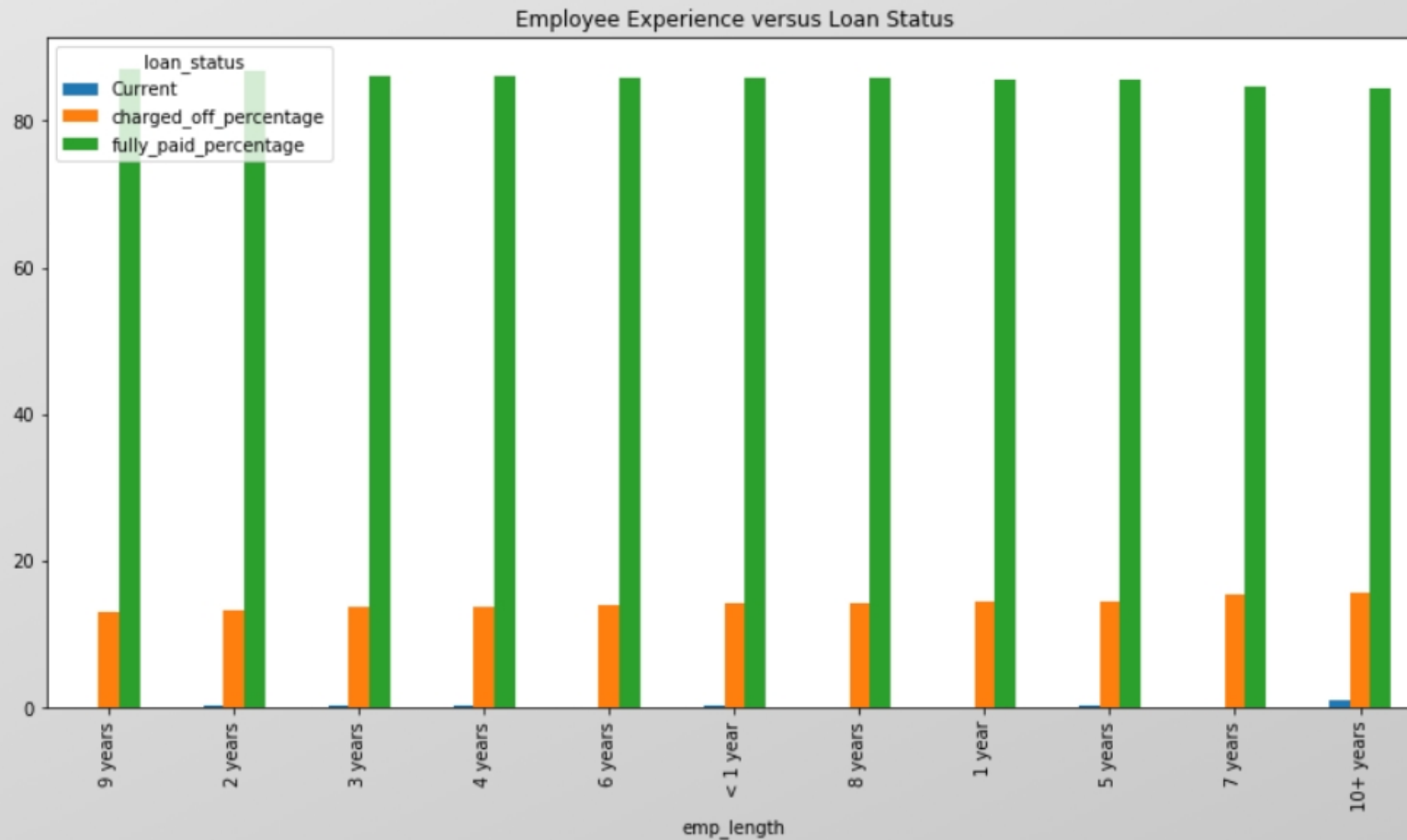
From the above plot we can see that 60 months term loans are more likely to get default.

Bivariate analysis two categorical variables loan status and dti



Significant correlation is not there between debt to income ratio and loan defaulting

Bivariate analysis two categorical variables loan status and employee length



Could not find any correlation between employee length and loan defaulting